

What is claimed is:

1. A system for detecting clusters in space and time using input data on occurrences of a phenomenon and characteristics at a plurality of locations and times comprising:

an expectation generation module determining expected occurrences of a phenomena at a plurality of locations and a plurality of times;

an occurrence modeling module determining actual occurrences of the phenomena at a plurality of locations and a plurality of times;

a search module searching the expected occurrences and the actual occurrences for a plurality of candidate solutions, wherein each solution is represented as a set of points in the three-dimensional space, and wherein each point corresponds to a location at a time;

a convex container module determining at least one solution corresponding to a selected convex container shape from the plurality of candidate solutions; and

a solution evaluation module determining a strength metric for each solution determined by the convex container module, the search module selecting a solution having a

desirable strength, wherein the solution having the desirable strength indicates a dominant cluster in the input data.

5 2. The system of claim 1, wherein the search module selects a strongest solution as determined by the solution evaluation module.

10 3. The system of claim 1, further comprising a cache module to save the solutions having the desired shape determined by the convex container module for previously examined sets of points.

15 4. The system of claim 1, wherein the input data on occurrences of a phenomenon comprise counts and times of the occurrences of the phenomenon at the locations in a time period.

20 5. The system of claim 1, wherein the input data on characteristics of the locations and times comprise of the populations subject to the occurrences of the phenomenon at the locations and times.

6. The system of claim 1, wherein the expectation generation model generates expected counts of occurrences at the locations and times using a Poisson model.

5 7. The system of claim 1, wherein the occurrence modeling module determines the occurrences as equal to the occurrences in the input data.

10 8. The system of claim 1, wherein the occurrence modeling module determines the occurrences at the locations and times based on their characteristics and a domain dependent model.

15 9. The system of claim 8, wherein the occurrences are determined from the population using a Poisson model.

20 10. The system of claim 1, wherein the search module considers candidate solutions represented as sets of points and utilizes the convex container module to determine solutions having the desired shape from the candidate solutions.

25 11. The system of claim 10, wherein the candidate solutions are initialized based on the input data.

12. The system of claim 11, wherein each initial candidate solution is singleton point.

13. The system of claim 10, wherein the search module
5 determines candidate solutions from solutions considered using an iterative process.

14. The system of claim 13, wherein candidate solutions are created from solutions considered based on the
10 chosen convex container shape.

15. The system of claim 1, wherein the convex container module determines a solution with minimum volume, given the selected convex container shape, that includes all
15 the points in a given candidate solution.

16. The system of claim 1, wherein the convex container module determines a solution with maximum volume, given the selected convex container shape, that excludes all
20 the points not in the given candidate solution.

17. The system of claim 1, wherein the convex container module determines a solution that maximizes a measure representing the equality between the set of points

in the given candidate solution and the set of points included in the solution, given the selected convex container shape.

5 18. The system of claim 1, wherein the solution evaluation module determines the strength metric based on all the points included in the solution and the expected occurrences determined by the expectation generation module and the occurrences determined by the occurrence modeling
10 module.

19. The system of claim 1, wherein the strength metric is based on the likelihood ratio using the spatial scan statistic.

15 20. A method for detecting clusters comprising:
receiving input data on occurrences of a phenomenon at locations and times and data on characteristics of the locations and times;

20 determining actual occurrences of the phenomenon at the three-dimensional space-time points according to the input data;

determining expected occurrences in the absence of any clustering of the phenomenon at the points according to the

characteristics of the occurrences using a domain dependent model for the phenomenon;

selecting a convex container shape for determining a cluster in the input data; and

5 determining a solution represented as a set of points that conforms to the selected convex container shape for a cluster with a desirable strength.

21. The method of claim 20, further comprising caching
10 solutions conforming to the selected convex container shape.

22. The method of claim 20, wherein determining expected occurrences further comprises generating expected counts of occurrences at the locations and times using a
15 model.

23. The method of claim 20, further comprising determining a strength of the cluster based on points included in the cluster, and the expected occurrences and
20 the actual occurrences.

24. A program storage device is provided readable by machine, tangibly embodying a program of instructions automatically executable by the machine to perform method

steps for determining a clusters, the method steps
comprising:

receiving input data on occurrences of a phenomenon at
locations and times and data on characteristics of the
5 locations and times;

determining actual occurrences of the phenomenon at the
three-dimensional space-time points according to the input
data;

determining expected occurrences in the absence of any
10 clustering of the phenomenon at the points according to the
characteristics of the occurrences using a domain dependent
model for the phenomenon;

selecting a convex container shape for determining a
cluster in the input data; and

15 determining a solution represented as a set of points
that conforms to the selected convex container shape for a
cluster with a desirable strength.

25 25. The method of claim 24, further comprising caching
solutions conforming to the selected convex container shape.

26. The method of claim 24, wherein determining
expected occurrences further comprises generating expected
counts of occurrences at the locations and times using a
25 model.

27. The method of claim 24, further comprising determining a strength of the cluster based on points included in the cluster, and the expected occurrences and the actual occurrences.